Articles

Side-Chain Determinants of β-Sheet Stability†

Daniel E. Otzen and Alan R. Fersht*

MRC Unit for Protein Function and Design and Cambridge Centre for Protein Engineering, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received December 14, 1994®

ABSTRACT: β-Sheet propensities of different amino acids depend on the context of both secondary and tertiary structure. In an attempt to establish general empirical relationships that determine this context dependence, we have determined the free energy of unfolding of a series of mutants at six positions in the β-sheet of chymotrypsin inhibitor 2 (CI2). We have generated the series Val→Ala→Gly and Val↔Thr at five positions, as well as the side-chain deletion Ile→Val at residue 49 and Ala→Gly at residue 77. In the series Val→Ala→Gly, the ranking order in terms of stability is Val > Ala > Gly at all positions. However, the change in free energy on deletion of methylene groups varies greatly. When Val and Thr are interchanged, the wild-type residue is always the more stable, but by a different amount at each position. We have attempted to rationalize the data by relating it to changes in solvent-accessible surface area, packing density, and statistically derived pseudo-energy functions that depend on ϕ,ψ angles. There is no significant correlation of the energies with any of the variables except with the pseudo-energy function, but the deviations from these values are large. We conclude that thermodynamic scales for β-sheet propensity are currently of insufficient precision for general design purposes, although they may be useful in special cases.

Proteins are stabilized by many different non-covalent intramolecular interactions. The same kind of mutation has different effects in different parts of a protein because of differences in the local environment, that is, a dependence on context. For example, the mutation of Ala to Gly in the solvent-exposed faces of the two α-helices of barnase destabilize it from 0.35 to 2.1 kcal/mol: the more buried the side chain of Ala, the more the destabilization (Serrano et al., 1992b,c). Further, there are even larger effects at the exposed ends (caps) of helices that depend on the exposure of hydrophilic groups (Serrano & Fersht, 1989, Serrano et al., 1992c). β-Sheet propensity also depends on the structural context (Minor & Kim, 1994a,b).

Several different types of correlation have been found with the changes in stabilization energy on the mutation of hydrophobic side chains to shorter ones. Mutations in the hydrophobic core of barnase (Serrano et al., 1992a) and chymotrypsin inhibitor 2 (Jackson et al., 1993) correlate well with the number of methyl and methylene groups that are within a sphere of 6 Å radius of the methylene groups that are removed on mutation. That is, for these buried groups, there is a correlation between stability and packing density. In contrast, Ala→Gly mutations in the solvent-exposed face of the helices of barnase show an excellent correlation between the change in stability and the change in solvent-accessible hydrophobic surface area on mutation (Serrano et al., 1992b). A further term has to be introduced when there are changes in the exposure of polar groups on mutation (Serrano et al., 1992b). Another correlation for the truncation of side chains in hydrophobic cores has been reported to be between the change in stability and the volume of the cavity so formed (Eriksson et al., 1992).

Thermodynamic scales for measuring the β-sheet-forming propensities of amino acids have been reported in a zinc-finger peptide (Kim & Berg, 1993) and the 56-residue immunoglobin-binding domain B1 from protein G (Smith et al., 1994; Minor & Kim, 1994a,b). The objective was to construct a scale by which the β-sheet propensities are expressed as free energy values relative to a reference amino acid ($\Delta\Delta G$).[1] In all cases, one particular highly solvent exposed residue with minimal local interactions in a β-sheet was replaced by all other amino acids. The different systems agree qualitatively, though the spans of $\Delta\Delta G$ values are 2.3 and 2.8 kcal/mol in the protein G system and 0.6 kcal/mol in the zinc-finger peptide system (for all substitutions of amino acids, barring Pro). The ranking orders for the β-sheet-forming propensities of the different amino acids differ in two different tertiary environments in the protein G system, namely, in a central strand (Minor & Kim, 1994a) and an edge strand (Minor & Kim, 1994b). Variable ranking orders were also found in a statistical analysis of the protein structural data base by Muñoz and Serrano (1994). By grouping the amino acids of each protein into different groups

† D.E.O. is supported by a predoctoral fellowship from the Danish Natural Science Research Council.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* April 15, 1995.

[1] Abbreviations: CI2, chymotrypsin inhibitor 2; ΔA_{HP}, change in solvent-accessible surface area of hydrophobic side-chain atoms (methyl and methylene groups) upon mutation; ΔA_{HB}, change in solvent-accessible surface area of polar atoms that require solvation upon mutation; $\Delta\Delta G_{calculated}$, change in stability upon mutation relative to wild type calculated from multiple linear regression; $\Delta\Delta G_{U-F}$, measured change in stability upon mutation relative to wild-type; $\Delta\Delta G_{int,X\rightarrow Y}$, intrinsic free energy change of mutating amino acid X to amino acid Y; GdnHCl, guanidinium hydrochloride; *r*, correlation coefficient; MES, 2-(*N*-morpholino)ethanesulfonic acid.

0006-2960/95/0434-5718$09.00/0

on the basis of their dihedral angles, Muñoz and Serrano generated pseudo-energy empirical scales, giving the propensities for different combinations of dihedral angles. The β-sheet propensities agree well with the average of the experimental scales reported so far. Their work suggested that the same amino acid can have different β-sheet propensities in different parts of a β-sheet, depending on its dihedral angles.

We present an analysis of 17 novel mutants with mutations in the β-sheet of chymotrypsin inhibitor 2 (CI2). The protein is small (64 residues in the truncated version used in this study), contains both an α-helix and a six-strand β-sheet, and lacks disulfide bridges. The protein unfolds according to a strict two-state mechanism (Jackson & Fersht, 1991). The six-strand β-sheet comprises residues 22−24, 30−32, 47−53, 64−70, 74−77, and 79−83. This β-sheet packs against the 13-residue α-helix to form the hydrophobic core. Mutations have been made at six different positions in the sheet, rather than concentrating on one or two specific sites. The side chains of the selected residues primarily interact with other side chains within the β-sheet and do not contact residues that are part of the hydrophobic core.

EXPERIMENTAL PROCEDURES

Materials. All chemicals used were as described by Jackson et al. (1993).

Methods. Naturally occurring CI2 is an 83-residue protein, where the first 19 residues are devoid of fixed structure. A truncated version lacking these first 19 residues was constructed (Jackson et al., 1993). Site-directed mutagenesis on the gene for this protein was carried out by inverse PCR (Innis & Gelfand, 1990). Mutants were identified by direct sequencing of the single-stranded DNA (Jackson et al., 1993), and the identity of the purified protein mutant was confirmed by electrospray mass spectroscopy. Expression and purification were carried out as described (Jackson et al., 1993).

Denaturation experiments with guanidinium hydrochloride were performed as described (Jackson et al., 1993). Briefly, the denaturation of CI2 is accompanied by a large increase in fluorescence when monitored by an excitation wavelength of 280 nm and an emission wavelength of 356 nm. The unfolding of each mutant was therefore followed by measuring the fluorescence of aliquots of the protein (typically 2.5 μM) at 25 °C in 50 mM MES, pH 6.3, at different concentrations of the denaturant guanidinium hydrochloride (GdnHCl).

Data Analysis. Data analysis was performed as described (Jackson et al., 1993). The free energy of unfolding of proteins in the presence of GdnHCl ($\Delta G_{U\text{-}F}{}^{D}$) is linearly related to the concentration of GdnHCl (Tanford, 1968):

$$\Delta G_{U\text{-}F}{}^{D} = \Delta G_{U\text{-}F}{}^{H_2O} - m_{U\text{-}F}[\text{GdnHCl}] \quad (1)$$

where $\Delta G_{U\text{-}F}{}^{H_2O}$ is the free energy of unfolding in water and $m_{U\text{-}F}$ is a constant. From this we can derive the equation

$$F = \{(\alpha_N + \beta_N[\text{GdnHCl}]) + \{(\alpha_U + \beta_U[\text{GdnHCl}]) \exp[m_{U\text{-}F}([\text{GdnHCl}]\} - [\text{GdnHCl}]^{50\%})/RT]\}/\{1 + \exp[m_{U\text{-}F}([\text{GdnHCl}] - [\text{GdnHCl}]^{50\%})/RT]\} \quad (2)$$

where F is the measured fluorescence; α_N and α_U are the intercepts, and β_N and β_U are the slopes, of the fluorescence base lines at low (N) and high (U) GdnHCl concentrations before and after the transition region of unfolding; $[\text{GdnHCl}]^{50\%}$ is the concentration of GdnHCl at which the protein is 50% denatured; and $m_{U\text{-}F}$ is the constant from eq 1. $\Delta\Delta G_{U\text{-}F}{}^{[G]50\%}$, the difference in stability between wild-type and mutant CI2 at a mean value of $[\text{GdnHCl}]^{50\%}$ for the two proteins, is calculated as

$$\Delta\Delta G_{U\text{-}F}{}^{[G]50\%} = \langle m_{U\text{-}F}\rangle \Delta[\text{GdnHCl}]^{50\%} \quad (3)$$

where $\langle m_{U\text{-}F}\rangle$ is the average value of $m_{U\text{-}F}$ for a large number of CI2 mutants (1.94 ± 0.025 kcal mol^{-2}) and $\Delta[\text{GdnHCl}]^{50\%}$ is the difference in $[\text{GdnHCl}]^{50\%}$ of the wild-type and mutant CI2. $\Delta\Delta G_{U\text{-}F}{}^{[G]50\%}$ has a smaller standard error than the difference in stability between wild-type and mutant CI2 in water ($\Delta\Delta G_{U\text{-}F}{}^{H_2O}$), due to shorter extrapolation (Jackson et al., 1993).

Structural Analysis. For structural analysis, the crystal structure of the double mutant Glu→Ala33/Glu→Ala34 (which we call the pseudo wild type) was used. The structures of mutant proteins were modeled by deleting the appropriate atoms in the coordinate file of the pseudo-wild-type crystal structure (Harpaz et al., 1994). Calculation of accessible surface area was carried out using a program written by Y. Harpaz (Centre for Protein Engineering, Cambridge), based on an algorithm employing a rolling sphere of radius 1.4 Å, similar to that of Lee and Richards (1971). The program produced a list of the solvent-accessible surface area of all atoms of the CI2 mutant. Surface area differences between wild type and mutant were calculated as follows. For each atom of the protein where the mutation altered the surface area compared to wild type, the area difference ΔA (in Å^2) between the mutant and the wild type was calculated by subtracting the area of the atom in the mutant atom from the area of the corresponding atom in the wild type. The change in solvent-accessible hydrophobic surface area (ΔA_{HP}) was then calculated by adding up ΔAs for all methyl(ene) carbon atoms in the protein, i.e., all carbon atoms except the carbonyl carbon. Similarly, the change in solvent-accessible surface area of polar atoms which require solvation (ΔA_{HB}) was calculated by adding up ΔAs for all carbonyl oxygens and amide nitrogens that do not hydrogen bond internally in the protein (Serrano et al., 1992b).

The packing density for a particular mutation was calculated by counting the number of methyl and methylene groups within 6 Å of each of the side-chain atoms removed by mutation (Serrano et al., 1992a).

The intrinsic secondary structure propensities of each amino acid at each mutated position in the β-sheet of CI2 were calculated according to Muñoz and Serrano (1994) as follows. The dihedral angles ϕ and ψ at each mutated position were calculated from the pseudo-wild-type crystal structure, in order to place the amino acid in one of the 400 squares into which the Ramachandran plot was divided (Muñoz & Serrano, 1994). $\Delta\Delta G_{int,X\to Y}$, the intrinsic free energy change of mutating amino acid X to amino acid Y, was calculated as

$$\Delta\Delta G_{int,X\to Y} = -RT[\ln(X_{\phi,\psi}/X_\Sigma) - \ln(Y_{\phi,\psi}/Y_\Sigma)] \quad (4)$$

where $X_{\phi,\psi}$ and $Y_{\phi,\psi}$ are the number of times that residues X

Table 1: Environment of Mutated Residues in the β-Sheet of CI2[a]

site of mutation in CI2	backbone hydrogen bonding	*B*-factor	dihedral angles (ϕ,ψ)[b]	side chains within 4.5 Å
Thr 22	O: Val 82 (N)	C^{β}: 19.5 $O^{\gamma 1}$: 21.0 $O^{\gamma 2}$: 20.5	−127.5, 6.4	V82, K21, E23 (all β-sheet)
Ile 49	N: Val 66 (O) O: Leu 68 (N)	C^{β}: 13.6 $C^{\gamma 1}$: 14.7 $C^{\gamma 2}$: 14.9 C^{δ}: 16.2	−116.3, 139.5	Q47, I48, V50, L51, R65, V66, R67 (all β-sheet)
Val 53	N: Val 70 (O)	C^{β}: 8.9 $C^{\gamma 1}$: 9.0 $C^{\gamma 2}$: 7.8	−58.3, 140.7	P52, G54, F69, V70, D71, A77, E78 (all β-sheet except G54 and D71)
Ala 77	N: Phe 69 (O)	C^{β}: 8.1	−103.1, −17.3	V53, F69, V70, D71, N75, I76, E78 (all β-sheet except D71)
Val 79	O: Phe 69 (N)[c]	C^{β}: 8.2 $C^{\gamma 1}$: 9.2 $C^{\gamma 2}$: 8.3	−61.1, 119.0	E23, W24, P25, V28, E78, P80, R81 (all β-sheet except P25 and V28)
Val 82	N: Thr 22 (O)	C^{β}: 8.8 $C^{\gamma 1}$: 8.6 $C^{\gamma 2}$: 9.7	−67.4, 139.0	K21, T22, W24, K43, Y61, R62, I63, R65, V66, R81, G83 (all β-sheet except K43, Y61, R62 and I63)

[a] Data based on pseudo-wild-type structure (Harpaz et al., 1994). [b] The (ϕ,ψ) angles of Val 53 and Ala 77 do not fall within the normal Ramachandran category of β-sheet (ϕ,ψ) angles. However, all the residues mutated in this study form hydrogen bonds to residues in other β-strands as indicated in the second column. [c] Via bridging water molecule.

and Y are found in the square in question, and X_{Σ} and Y_{Σ} are the total number of residues of this kind present in the data base.

RESULTS

The 17 mutants used for our structural analysis were constructed at six different positions in the β-sheet of chymotrypsin inhibitor 2 (CI2). The side chains of these residues interact predominantly with other residues within the β-sheet (Table 1). The mutations comprise deletions of methylene groups, as well as isosteric replacements of Val with Thr and vice versa. In all cases (except for the mutation Ile→Val 49), removal of methylene group leads to a loss in stability (Table 2). Mutations at all sites retain the β-sheet propensity ranking Val > Ala > Gly, but the difference in free energy of unfolding between each type amino acid varies greatly (Table 4). When Val and Thr are exchanged, the wild-type residue is in all cases the stabler of the two. All the mutations are accompanied by changes in solvent-accessible surface area and packing density as determined by modeling of mutants (Table 3). We examined, therefore, the stability data for significant correlations with structural changes.

Stability and Packing Density. Packing density is not an important factor in determining β-sheet stability. The correlation coefficient between $\Delta\Delta G_{U\text{-}F}$ and packing density is very low, whether all hydrophobic mutants are considered together (the correlation coefficient $r < 0.1$ for 11 points; Figure 1A) or in two different classes ($r = 0.3$ for both Ala→Gly and Val→Ala mutants).

Stability and Solvent-Accessible Surface Areas of Hydrophobic Truncation Mutants. For hydrophobic truncation mutations in the β-sheet of CI2 (i.e., not involving Val↔Thr mutations), the correlation between $\Delta\Delta G_{U\text{-}F}$ and the change in solvent-accessible hydrophobic surface area is very poor: $r = 0.45$ for 12 points (Figure 1B).

When ΔA_{HP} and ΔA_{HB} of hydrophobic truncation mutations are combined by multiple linear regression and the measured $\Delta\Delta G_{U\text{-}F}$ is plotted against $\Delta\Delta G_{calculated}$ ($\Delta\Delta G$ of unfolding is calculated for each mutation according to the equation derived from multiple linear regression), the correlation remains poor: $r = 0.5$ for 12 points (Figure 1C).

Table 2: Equilibrium Denaturation Parameters for CI2 Mutants

mutant	m_{U-F}[a]	[GdnHCl]$^{50\%}$ [b] (kcal/mol)	$\Delta\Delta G_{U-F}^{[G]50\%}$ [c] (kcal/mol)
wild type	1.90 ± 0.05	4.00 ± 0.01	0.00
Thr→Val22	1.67 ± 0.08	3.83 ± 0.03	0.32 ± 0.07
Thr→Ala22	1.83 ± 0.09	3.56 ± 0.02	0.85 ± 0.05
Thr→Gly22	1.95 ± 0.12	3.40 ± 0.03	1.16 ± 0.06
Ile→Val49	1.77 ± 0.04	4.04 ± 0.04	−0.08 ± 0.08
Ile→Ala49	2.22 ± 0.09	2.91 ± 0.02	2.12 ± 0.06
Ile→Gly49	2.13 ± 0.08	2.19 ± 0.01	3.52 ± 0.07
Ile→Thr49	1.97 ± 0.07	3.31 ± 0.02	1.34 ± 0.04
Val→Thr53	1.65 ± 0.07	3.47 ± 0.02	1.03 ± 0.05
Val→Ala53	1.70 ± 0.15	3.67 ± 0.05	0.64 ± 0.11
Val→Gly53	2.00 ± 0.08	2.75 ± 0.01	2.43 ± 0.05
Ala→Gly77	2.19 ± 0.22	3.03 ± 0.04	1.88 ± 0.08
Val→Thr79	1.63 ± 0.14	3.80 ± 0.06	0.38 ± 0.11
Val→Ala79	1.92 ± 0.11	3.22 ± 0.03	1.51 ± 0.06
Val→Gly79	2.61 ± 0.11	2.33 ± 0.01	3.24 ± 0.06
Val→Thr82	1.79 ± 0.13	3.41 ± 0.03	1.15 ± 0.07
Val→Ala82	1.95 ± 0.14	3.25 ± 0.03	1.45 ± 0.07
Val→Gly82	2.10 ± 0.06	2.20 ± 0.01	3.50 ± 0.07

[a] m_{U-F} is the dependence of the free energy of unfolding on denaturant concentration [D]; i.e., $\Delta G_{U-F}^{D} = \Delta G_{U-F}^{H_2O} - m_{U-F}[D]$. The denaturant is GdnHCl. [b] [GdnHCl]$^{50\%}$ is the concentration of GdnHCl at which the protein is 50% unfolded. [c] $\Delta\Delta G_{U-F}^{[G]50\%}$ is the difference in free energy of unfolding between the mutant and wild type at a mean value of [GdnHCl]$^{50\%}$ for the two proteins.

Separate examinations of Ala→Gly and Val→Ala mutants also reveal poor correlations with hydrophobic surface area changes ($r = 0.5$ for six and five points, respectively) with large standard errors for linear fittings. Thus slopes (for $\Delta\Delta G_{U\text{-}F}$ vs $\Delta\Delta A_{HP}$) are 0.04 ± 0.03 and 0.07 ± 0.06, respectively. Mutliple linear regression performed on each class of mutants, which includes terms for accessible surface areas for hydrogen bonding to solvent, as well as $\Delta\Delta A_{HP}$, leads to better correlations between $\Delta\Delta G_{U\text{-}F}$ and $\Delta\Delta G_{calculated}$

Table 3: Surface Area Changes upon Mutation in the β-Sheet of CI2[a]

position in CI2 sequence	increase in total solvent-accessible hydrophobic surface area (ΔA_{HP}) (Å^2)	decrease in solvent-accessible surface area of hydrophilic groups requiring solvation (ΔA_{HB}) (Å^2)	no. of methyl(ene) groups within 6.0 Å of carbon atom removed by mutation
Ala→Gly mutations			
22	8.0	6.4	19
49	−7.3	4.3	16
53	8.2	2.0	15
77	−1.2	2.1	18
79	2.8	5.2	16
82	−14.7	0.9	19
Val→Ala mutations			
22	6.7	1.5	21
49	4.2	11.3	22
53	10.5	9.5	23
77	2.9	3.9	31
82	−4.2	1.5	35
Ile→Val mutation			
49	2.8	3.8	10

[a] All structural data are based on the pseudo-wild-type crystal structure (Harpaz et al., 1994).

Table 4: Comparison of Different Thermodynamic Scales for β-Sheet Propensities

protein	$\Delta\Delta G_{U-F}$ (kcal/mol)			
	Ile→Val	Val→Ala	Ala→Gly	Val→Thr
zinc-finger peptide CP1[a]	0.03	0.18	0.35	0.05
B1 domain of IgG binding protein G[b] (central strand)	0.31	0.94	1.21	−0.69
B1 domain of IgG binding protein G[c] (central strand)	0.18	0.82	1.2	−0.28
B1 domain of IgG binding protein G[d] (edge strand)	−0.15	0.17	0.83	−0.66
CI2 residue 22	*e*	0.52 ± 0.08	0.31 ± 0.08	−0.32 ± 0.07
CI2 residue 49	−0.08 ± 0.08	2.20 ± 0.10	1.40 ± 0.09	1.34 ± 0.04
CI2 residue 53	*e*	0.64 ± 0.11	1.80 ± 0.12	1.03 ± 0.05
CI2 residue 77	*e*	*e*	1.88 ± 0.08	*e*
CI2 residue 79	*e*	1.51 ± 0.06	1.72 ± 0.08	0.38 ± 0.11
CI2 residue 82	*e*	1.45 ± 0.07	2.05 ± 0.09	1.15 ± 0.07

[a] Kim and Berg (1993). [b] Smith et al. (1994). [c] Minor and Kim (1994a). [d] Minor and Kim (1994b). [e] Mutant not constructed.

(0.82 and 0.9 for the Ala→Gly and Val→Ala mutants, respectively). These apparently high correlation coefficients do not mean that the correlation is significant, but arise because there are too few data points. There are also large standard errors for each term and a great disparity between the actual regression parameters for each class. For Ala→Gly mutants (Figure 1D), the parameters are

$$\Delta\Delta G_{U-F} = 2.3(\pm 0.4) - 0.01(\pm 0.03)\Delta A_{HP} + 0.22(\pm 0.11)\Delta A_{HB}\ \text{(kcal/mol)} \quad (5)$$

whereas the parameters for Val→Ala mutants are

$$\Delta\Delta G_{U-F} = 1.0(\pm 0.3) - 0.12(\pm 0.05)\Delta A_{HP} - 0.13(\pm 0.05)\Delta A_{HB}\ \text{(kcal/mol)} \quad (6)$$

The inclusion of a term for hydrogen bonding to solvent improves the correlation between $\Delta\Delta G_{U-F}$ and $\Delta\Delta G_{calculated}$ for the two classes of mutations, Val→Ala and Ala→Gly, when they are considered separately, but not when they are considered together. Indeed, a plot of ΔA_{HB} against $\Delta\Delta G_{U-F}$ for both mutant classes gives a correlation coefficient of less than 0.1

Inclusion of the packing parameter, i.e., the number of methyl(ene) groups within 6 Å of the mutated atoms, in combination with ΔA_{HP} and ΔA_{HB}, gives no improvement to the overall fit between $\Delta\Delta G_{U-F}$ and $\Delta\Delta G_{calculated}$ (the latter calculated for each hydrophobic truncation mutant on the basis of the multiple regression fit). Again, the overall correlation between is improved when the two classes of mutants, Ala→Gly and Val→Ala, are considered separately. Linear fits of $\Delta\Delta G_{calculated}$ versus measured $\Delta\Delta G_{U-F}$ for the different groups of mutants give correlation coefficients and confidence limits very similar to those observed in the absence of the packing parameter.

Val↔Thr Mutations. The isosteric interchange between Val and Thr does not alter the overall surface area of the protein (provided no structural rearrangements occur), but it changes the distribution of hydrophobic and hydrophilic surface area. This also affects the stability of the unfolded state, where a Thr side chain interacts more favorably with solvent than a Val side chain. Given that the unfolded state of CI2 is by good approximation a random coil (Jackson et al., 1991), Val↔Thr mutations in different parts of the protein affect the stability of the unfolded state equally. We will, therefore, confine ourselves to examining changes in the folded state.

Since the hydroxyl group can replace either $C^{\gamma 1}$ or $C^{\gamma 2}$, we will consider the sum of the solvent-accessible surface areas of $C^{\gamma 1}$ (or $O^{\gamma 1}$ in the case of Thr22) and $C^{\gamma 2}$, and the packing around both $C^{\gamma 1}$ and $C^{\gamma 2}$. The mutation Thr→Val22 is incorporated with the rest of the Val to Thr mutations by reversing the sign of its $\Delta\Delta G_{U-F}$ value. There is no correlation between stability changes and surface area

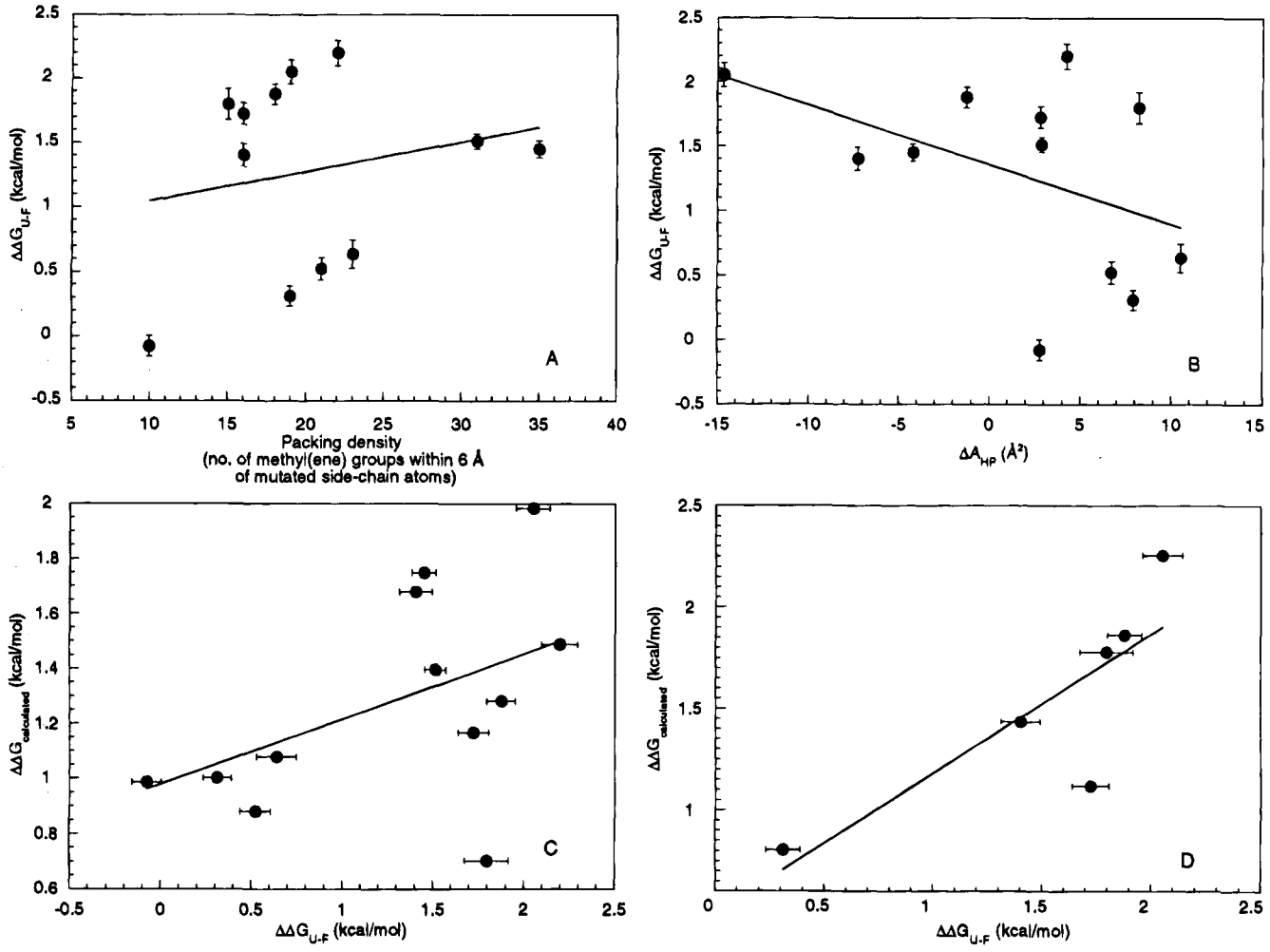

FIGURE 1: Correlations of $\Delta\Delta G_{U\text{-}F}$ for mutations in the β-sheet of CI2 with different structural parameters. (A) Correlation of $\Delta\Delta G_{U\text{-}F}$ for hydrophobic mutants with packing density. (B) Correlation of $\Delta\Delta G_{U\text{-}F}$ for hydrophobic mutants with change in solvent-accessible hydrophobic surface area, ΔA_{HP}. (C) Correlation of $\Delta\Delta G_{calculated}$ with $\Delta\Delta G_{U\text{-}F}$ for hydrophobic mutants based on a multiple linear regression analysis, involving both ΔA_{HP} and ΔA_{HB}. (D) Correlation of $\Delta\Delta G_{calculated}$ with $\Delta\Delta G_{U\text{-}F}$ for mutations Ala→Gly. $\Delta\Delta G_{calculated}$ is calculated from the equation derived from a mutliple linear regression analysis between $\Delta\Delta G_{U\text{-}F}$ and the parameters ΔA_{HP}, ΔA_{HB}, and packing.

changes alone (r = 0.1, 5 points). When the packing parameter is included for multiple regression analysis, the correlation r between $\Delta\Delta G_{calculated}$ and $\Delta\Delta G_{U\text{-}F}$ increases to 0.71 for 5 points, but this is still below the 90% confidence limit. For comparison, the correlation between $\Delta\Delta G_{calculated}$ and $\Delta\Delta G_{U\text{-}F}$, when only surface and packing changes around $C^{\gamma 1}$ are considered, is 0.1.

Stability and Statistical Parameters (Pseudo-Energy Scales). The pseudo-energy empirical scales constructed by Muñoz and Serrano (1994) enable us to calculate $\Delta\Delta G_{int,X\rightarrow Y}$, the intrinsic free energy change of mutating residue X to Y, which depends only on the dihedral angles ϕ and ψ at the position of mutation. In this way, it is possible to take into consideration local structural differences within elements of secondary structure, which are ignored in most statistical analyses [e.g., Fasman (1989)].

Figure 2 shows the plot of $\Delta\Delta G_{U\text{-}F}$ versus $\Delta\Delta G_{int,X\rightarrow Y}$ for all mutants except Ile→Val49 (for which no statistical Ramachandran matrix was available). The linear fit is

$$\Delta\Delta G_{U\text{-}F} = 0.98(\pm 0.14) + 0.91(\pm 0.23)^{*}\Delta\Delta G_{int,X\rightarrow Y}\ \text{(kcal/mol)} \quad (7)$$

(r = 0.72, 16 points) (standard errors in parentheses). Although the errors are large, this correlation is signifciant above the 99% confidence limit. Correlations for separate groups of mutants are lower, and all are below the 90% confidence limit. Multiple linear regression analysis incorporating changes in exposed surface area and packing density did not significantly improve this correlation (data not shown).

DISCUSSION

We have measured the stability changes for 17 mutants in the β-sheet of the 64-residue protein CI2 in an attempt to obtain simple empirical rules correlating structure and stability of β-sheets. All the mutations remove interactions without replacing them with new ones, except to change the solvent-accessible surface area and (in the case of Val↔Thr mutations) replace a nonpolar atom with a polar one.

In line with conclusions from Minor and Kim (1994b), our studies suggest that a β-sheet propensity scale *per se* is not of much practical use in terms of either ranking or range of values. The propensities vary depending on both the protein and the site of mutation in the protein (Table 4). In CI2, Thr is more β-sheet-stabilizing than Ala at positions 22, 49, and 82 but less stabilizing at position 53. Similarly, Val is more stabilizing than Thr at positions 49, 53, 79, and 82 but less stabilizing at position 22. The free energy of

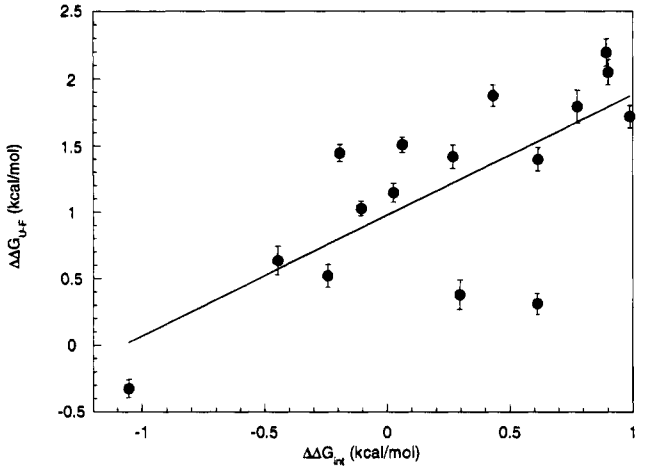

FIGURE 2: Correlation of $\Delta\Delta G_{U\text{-}F}$ with $\Delta\Delta G_{int,X\rightarrow Y}$ for all mutations in the β-sheet of CI2 except Ile→Val49.

destabilization within each class of mutation also varies. There is support, however, for the rank order Val > Ala > Gly, which form a straightforward hydrophobic series. This rank order is also observed in a model system involving a heteroaromatic diacylaminoepindolidione template (Kemp, 1990).

β-Sheets are formed by aligning residues distant in sequence. Therefore, interactions with surrounding residues are likely to play a larger role in determining β-sheet propensities at individual positions than is observed for α-helices. This emphasizes the importance of analyzing the structural consequences of such mutations. ΔA_{HB}, ΔA_{HP}, and packing density are all good candidates for determinants of β-sheet stability. The change in solvent-accessible nonpolar and polar surface area upon mutation is known to be important in determining helical stability (Serrano et al., 1992b,c). Further, hydrophobic clustering of side-chain atoms has been shown to be essential for β-sheet nucleation in a system employing dibenzofuran-based amino acids as abiotic β-hairpin analogs (Tsang et al., 1994).

However, packing density in the β-sheet does not correlate significantly with $\Delta\Delta G_{U\text{-}F}$. Nor is there any correlation between $\Delta\Delta G_{U\text{-}F}$ and ΔA_{HB}, which describes the change in solvent-accessible surface area of atoms that can hydrogen bond to solvent. The barnase helices and the CI2 β-sheet have similar slopes for plots of $\Delta\Delta G_{U\text{-}F}$ vs ΔA_{HP} (55 and 46 cal/mol/Å^2, respectively), but the correlation in the case of the β-sheet is not significant above the 90% limit.

Analysis of Val↔Thr mutations in the absence of detailed knowledge of mutant structures is complicated because either of the C^γ atoms can be replaced with oxygen when Val is mutated to Thr. Attempts to correlate stability changes with changes in total surface area and packing of both $C^{\gamma 1}$ and $C^{\gamma 2}$ atoms do not yield significant correlations. Previous work (Blaber et al., 1993) has shown that substitution of largely exposed Val by Thr results in ≤0.5 kcal/mol destabilization, while replacement of completely buried Val residues destabilizes the protein by 1−3 kcal/mol. This agrees with the results of mutations at positions 79 (destabilized by 0.4 kcal/mol; $C^{\gamma 2}$ exposes 31 Å^2 to solvent) and 82 (destabilized by 1.2 kcal/mol; both C^γ atoms are almost completely buried) in CI2. Val→Thr22 stabilizes the protein by 0.3 kcal/mol, and $O^{\gamma 1}$ at Thr22 has 26 Å^2 exposed. The relatively large destabilizations of Val→Thr53 and Ile→Thr49 (1.0 and 1.3 kcal/mol, respectively) can be rationalized in a rather facile manner by suggesting that the highly buried residues $C^{\gamma 1}$ (4 Å^2 exposed) in Val53 and $C^{\gamma 2}$ (5 Å^2 exposed) in Ile49 are the ones replaced by a hydroxyl atom. However, this rationalization lacks predictive power.

The best correlation with $\Delta\Delta G_{U\text{-}F}$ for all mutants in the β-sheet of CI2 is obtained by considering $\Delta\Delta G_{int,X\rightarrow Y}$, the intrinsic change in free energy upon mutation due to different secondary structural propensities of each amino acid (Muñoz & Serrano, 1994). This correlation is significant above the 99% limit, better than all the other correlations we have obtained. However, the errors for the linear fit between $\Delta\Delta G_{U\text{-}F}$ and $\Delta\Delta G_{int,X\rightarrow Y}$ are large (eq 4). Conceivably the correlation could be improved by increasing the resolution of the Ramachandran ϕ,ψ matrix, which is arbitrarily divided into 400 squares (Muñoz & Serrano, 1994). The correlation is not improved by including changes in solvent-exposed surface area or packing density, contrary to what is found for a series of Ala→Gly mutations in the protein CheY (E. Lopez & L. Serrano, unpublished data).

It has been argued that the side chain of a particular residue favors particular dihedral angles because these angles may facilitate solvation of the side chain or optimize packing (Bai & Englander, 1994). In other words, packing and solvation should determine the favored dihedral angles. In fact, we find that dihedral angles, and not surface area changes or packing density, are the best parameters for determining site-specific β-sheet propensity among the structural parameters tested in our study. But the large errors in the correlation argue that determination of such propensity with a higher degree of accuracy must also take into account more subtle structural effects, such as the size, water accessibility, and water occupancy of the cavity formed by the mutation and the exposure of hydrophobic and hydrophilic atoms within this cavity. This will require high-resolution structures of all mutants, illustrating the limitations which hamper present protein design in its attempts to predict the energetic consequences of even relatively simple non-disruptive mutations.

ACKNOWLEDGMENT

We are grateful to Dr. L. Serrano for communicating work to us prior to publication. Work presented in this paper was inspired by discussions at the International Summer School on Protein Structure, Function and Design on Spetsai, Greece, in August−September 1993.

REFERENCES

Bai, Y., & Englander, S. W. (1994) *Proteins: Struct., Funct., Genet. 18*, 262−266.

Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Dirk, W. H., & Matthews, B. W. (1993) *Biochemistry 32*, 11363−11373.

Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178−183.

Fasman, G. D. (1989) In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) Plenum, New York.

Fersht, A. R., & Serrano, L. S. (1993) *Curr. Opin. Struct. Biol. 3*, 75−83.

Harpaz, Y., elMasry, N., Fersht, A. R., & Henrick, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 311−315.

Innis, M. A., & Gelfand, D. H. (1990) in *PCR protocols* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J., Eds.) pp 3−12, Academic Press, New York.
Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., & Fersht, A. R. (1993) *Biochemistry 32*, 11259−11269.
Kemp, D. S. (1990) *Trends Biotechnol. 8*, 249−255.
Kim, C. A., & Berg, J. M. (1993) *Nature 362*, 267−270.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
Ludvigsen, S., Shen, H., Kjær, M., Madsen, J. C., & Poulsen, F. M. (1991) *J. Mol. Biol. 22*, 621−635.
McPhalen, C. A., & James, M. N. G. (1987) *Biochemistry 26*, 261-269.
Minor, D. L., & Kim, P. S. (1994a) *Nature 367*, 660−663.
Minor, D. L., & Kim, P. S. (1994b) *Nature 371*, 264−267.
Muñoz, V., & Serrano, L. (1994) *Proteins: Struct., Funct., Genet. 20*, 301−311.
Serrano, L., & Fersht, A. R. (1989) *Nature 342*, 296−299.
Serrano, L., Kellis, J. T., Cann, P., Matouschek, A., & Fersht, A. R. (1992a) *J. Mol. Biol. 224*, 783−804.
Serrano, L., Neira, J. L., Sancho, J., & Fersht, A. R. (1992b) *Nature 356*, 453−455.
Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. R. (1992c) *J. Mol. Biol. 227*, 544−559.
Smith, C. K., Withka, J. M., & Regan, L. (1994) *Biochemistry 33*, 5510−5517.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121−282.
Tsang, K. Y., Diaz, H., Graciani, N., & Kelly, J. W. (1994) *J. Am. Chem. Soc. 116*, 3988−4005.

BI942873+